Predicting Loan Default: Model Comparison and Evaluation

Analyzing and Comparing Logistic Regression, Decision Tree, and Random Forest Models



Project Overview

Objective

Predict loan default

Data

Customer and loan data

Models Used

- Logistic Regression
- Decision Tree
- Random Forest

Data Preprocessing

Missing Values

Used unknown link_numeric to convert data types and handle non-numeric data

Normalization

Standardized data using StandardScaler

Class Imbalance

Addressed using SMOTE

Model Training

Training Process

- Split data into training and testing sets
- Applied SMOTE on training set
- Trained models on resampled data

Hyperparameter Tuning

Used GridSearchCV



Model Comparison

1 Accuracy

Proportion of correct predictions

Precision

Correct positive predictions

3 Recall

Correct positive predictions

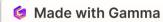
4 F1-Score

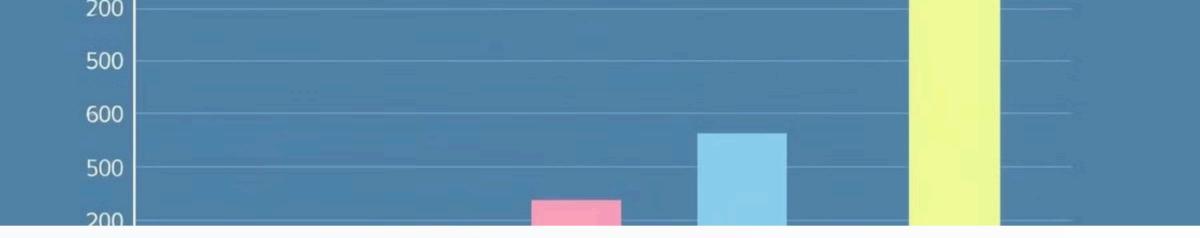
Harmonic mean of precision and recall

Nul Accuracy	Rocel Tecuracy	Put (F-sore)	Round Tecurarcy	Goff (F1-score)	
90.1%	54.55	\$8.195	74.99	70.0.35	
3.9%	7.721	3.135	27.22	76.0.00	
8.84	3.195	4.166	74.49	35.0.05	
2.3%	3.246	2.168	37.49	30.0.33	
3.4%	3.196	2.65	15.55	36.0.35	
3.8%	3.735	1.115	29.06	39.0.40	
0.8%	3.286	1.164	27.75	35.0.17	
11.8%	1346	2.65	15.89	10.0.31	
11.2%	4.988	1.106	14.99	15.0.02	
3.9%	7.44	1.80	8.90	-	
	Accuracy 90.1% 3.9% 8.84 2.3% 3.4% 3.8% 0.8% 11.8%	Accuracy Tecuracy 90.1% 54.55 3.9% 7.721 8.84 3.195 2.3% 3.246 3.4% 3.196 3.8% 3.735 0.8% 3.286 11.8% 1346 11.2% 4.988	Accuracy Tecuracy (F-sore) 90.1% 54.55 \$8.195 3.9% 7.721 3.135 8.84 3.195 4.166 2.3% 3.246 2.169 3.4% 3.196 2.65 3.8% 3.735 1.115 0.8% 3.286 1.164 11.8% 1346 2.65 11.2% 4.988 1.106	Accurary Tecuracy (F-sore) Tecurarcy 90.1% 54.55 \$8.195 74.99 3.9% 7.721 3.135 27.22 8.84 3.195 4.166 74.49 2.3% 3.246 2.168 37.49 3.4% 3.196 2.65 15.55 3.8% 3.735 1.115 29.06 0.8% 3.286 1.164 27.75 11.8% 1346 2.65 15.89 11.2% 4.988 1.106 14.99	

Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regressio n	0.753	0.917	0.753	0.827
Decision Tree	0.810	0.907	0.844	0.875
Random Forest	0.845	0.918	0.881	0.899





Results and Discussion

1 Random Forest
Best model

Decision Tree
Competitive performance

3 Logistic Regression
Room for improvement



Conclusion

Best Model

Random Forest

Key Takeaways

- High accuracy
- Balanced performance